Variational Image-Based Rendering with Gradient Constraints

Grégoire Nieto, Frédéric Devernay, James L. Crowley

VARIATIONAL IMAGE-BASED RENDERING WITH GRADIENT CONSTRAINTS

Grégoire Nieto[1], *Frédéric Devernay*[1], *James Crowley*[2]

[1]Laboratoire Jean Kuntzmann, Univ. Grenoble Alpes & INRIA, France.
[2]Laboratoire d'Informatique de Grenoble, Univ. Grenoble Alpes & INRIA, France.

ABSTRACT

Multi-view image-based rendering consists in generating a novel view of a scene from a set of source views. In general, this works by first doing a coarse 3D reconstruction of the scene, and then using this reconstruction to establish correspondences between source and target views, followed by blending the warped views to get the final image. Unfortunately, discontinuities in the blending weights, due to scene geometry or camera placement, result in artifacts in the target view. In this paper, we show how to avoid these artifacts by imposing additional constraints on the image gradients of the novel view. We propose a variational framework in which an energy functional is derived and optimized by iteratively solving a linear system. We demonstrate this method on several structured and unstructured multi-view datasets, and show that it numerically outperforms state-of-the-art methods, and eliminates artifacts that result from visibility discontinuities.

Index Terms— Image-Based Rendering, Computational Photography, Computer Graphics

1. INTRODUCTION

Multi-view image-based rendering consists in generating a novel view of a scene from a set of source views. In general, this works by first doing a coarse 3D reconstruction of the scene, called a *geometric proxy*, then using this reconstruction to establish correspondences between source and target views(Figure 1 (a)), followed by blending the warped views to obtain the final image. Recent work by Pujades *et al* [24] proposed a Bayesian formulation of the image-based rendering problem, building on previous work by Wanner and Goldluecke [11, 29]. They showed that the weight of each source image in the novel view could be formally deduced from the camera properties, image content, and accuracy of the geometric proxy, leading to a formalization of the heuristic blending weights proposed initially by Buehler *et al* [3]. Most of the "desirable properties that [...] an ideal image-based rendering algorithm should have" [3] were thus given a formal explanation, except for the *continuity* property. Therefore, discontinuities in the source image weights, which are mainly due to scene geometry or camera placement, result in artifacts in the target view.

In this work, we show that a way to avoid these artifacts is to impose additional constraints on the image gradients of the novel view. These constraints come from a simple observation: an image contour in the target image should also be present in source images where this part of the scene is visible. An energy functional similar to the one in Pujades *et al*is derived, which is composed of the usual *data* term and *smoothness* term, but the data term has an additional term which takes into account these gradient constraints. The resulting energy is optimized in a variational framework, by iteratively solving a linear system.

Our method is composed of two stages: a 3D reconstruction pipeline (section 4.1) followed by the target view rendering (section 3). The depth maps we obtain from multi-view stereo are used to compute a warp function from each input image to the novel view. The target view is computed via a novel variational formulation of the image-based rendering problem. We demonstrate the method on several unstructured multi-view datasets (section 4) and show that not only it numerically outperforms state-of-the-art methods (section 4.2), but also it eliminates artifacts that originated from visibility discontinuities. We conclude that taking into account both intensities and gradients in image-based rendering methods offers an elegant solution to enforcing the *continuity* property initially devised by Buehler *et al*.

2. PREVIOUS WORK

Image-Based Rendering (IBR) has been extensively reviewed by Shum *et al* [25]. Most state-of-the-art [15, 26, 17, 18, 5] approaches use a coarse 3D reconstruction of the scene, called a *geometric proxy*, which may have various degrees of accuracy. Ortiz-Cayon *et al* [22] choose to over-segment the image and compute the quality of several IBR algorithms on each super-pixel. Then the output of the best IBR algorithm for each super-pixel is picked. These algorithms are largely inspired by the Unstructured Lumigraph [3], that performs a blending of the k-nearest views, weighted by angles and distances to the target view, thus guarantying a smooth camera blending field. The continuity of the resulting blend in image domain is ensured by enforcing spatial smoothness on

Thanks to DGA for funding

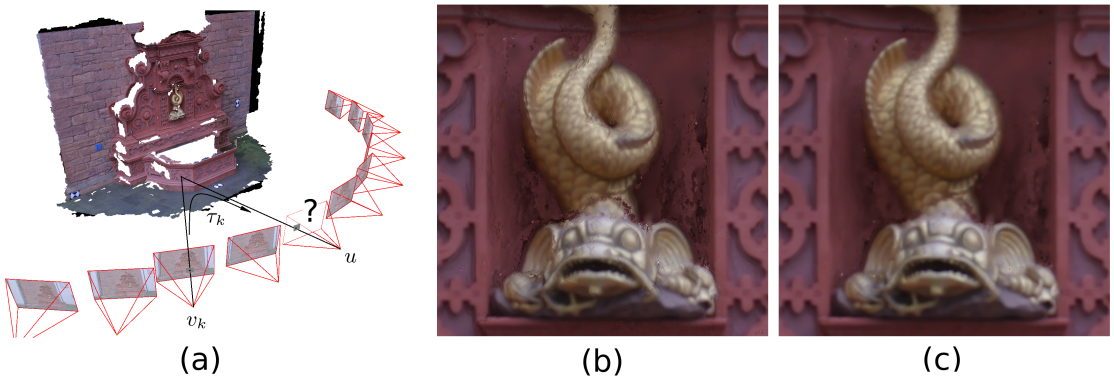


Fig. 1: (a) We coarsely estimate the geometry of the scene to register input images. The warp functions τ_k link the input images v_k to the target view u. (b) The target view u is reconstructed by blending the source images in intensities. (c) Our method consists in appending constraints on the gradient of the solution; we show that it is equivalent to a Laplacian blending and that it removes high frequency blending artifacts.

these weights, but temporal artifacts may still occur if the contributing cameras are too sparse. Davis *et al* [6] propose a viewpoint-subdivision rendering technique to create a better blending field. Nevertheless, blending weights are still heuristics and the choice of cameras for rendering is arbitrary.

Disposing of heuristics and tunable parameters is a key objective in [29], who propose a physics-based Bayesian formulation of image-based rendering. The weight of each input view in the final blending is automatically deduced from the mathematical equations by deriving an energy functional. Pujades *et al* [24] went further by integrating geometric uncertainty in the Bayesian formalism. They then obtain new weights that favor cameras satisfying both *epipole consistency* and *minimal angular deviation*, two principles stated by Buehler *et al* [3] to describe the ideal IBR algorithm. They were unable, however, to provide a formal derivation that leads to the *continuity* principle, especially near the borders of each camera view field. We show that introducing an additional term in the energy functional, not only constrains the image intensities but also constrains the image gradients, providing an elegant solution to the *continuity* principle.

The key idea of high-quality rendering is that the quality of the image solution often relies on the constraints put on the search space. As a consequence, finding the right regularization or image prior has been widely researched in order to obtain high-quality images. The main contribution of Fitzgibbon *et al* [7] is the use of texture priors computed from a large database of patches to constrain the solution, independently from the input scene data. This idea was recently extended by Flynn *et al* [8] who perform new view synthesis via a deep network architecture, trained by a huge database of real-world image sets. In contrast our method does not rest upon any strong prior knowledge about the new view to synthesize, but rather makes better use of the data provided by the input views to add more constraints on the solution. Therefore our algorithm does not require large numbers of picture sets from the world's imagery to create a high-quality images.

Image fusion in the gradient domain has received much interest in recent years beginning with the seminal paper of Perez *et al* [23], for applications such as image editing [19], inpainting [16] or image stitching [30]. The closest work to ours for image-based rendering is probably [15], who generate a new point of view by reconstructing the gradients in the target view, followed by an integration of these gradients on the GPU to recover the final image. However, their method is limited to interpolating between two views, and they neither address the generic case where the input views may have very different viewpoints, fields of view, and resolution, nor produce super-resolved images. Our method addresses these issues and proposes a more generic framework for multi-view image-based rendering.

3. VARIATIONAL IMAGE-BASED RENDERING

Our goal is to synthesize an optimal novel image $u : \Gamma \subset \mathbb{R}^2 \to \mathbb{R}$ at the target viewpoint from the input images $v_k : \Omega_k \subset \mathbb{R}^2 \to \mathbb{R}$. For the sake of simplicity, image values are taken as scalars, but this easily generalizes to color images $v_k : \Omega_k \to \mathbb{R}^3$. In all our experiments we process the images in the RGB color space. Source images are registered via the warps τ_k that transform any point $\mathbf{x}_m = (x_m, y_m)^\intercal$ of the input view k into the corresponding point $\mathbf{x}_p = (x_p, y_p)^\intercal$ in the target view:

$$\begin{array}{rccc} \tau_k : & \Omega_k & \to & \Gamma \\ & \mathbf{x}_m & \mapsto & \mathbf{x}_p \end{array} \tag{1}$$

3.1. Image Formation Model

As commonly assumed in the super-resolution literature [1, 14], we consider the intensity value $v_k(\mathbf{x}_m)$ at a point $\mathbf{x}_m$

in the low-resolution observed image k to be the convolution of values in the super-resolved image with the point-spread function (PSF) b. Given an ideal super-resolved image u at the position of the target view defined over Γ, and a warp τ_k that maps points from Ω_k (low resolution domain) to Γ (high resolution), if we discard for now the visibility effects, the intensity in the observed image k can be written as:

$$v_k(\mathbf{x}_m) = \int_{\Omega_k} u \circ \tau_k(\mathbf{x})\, b(\mathbf{x} - \mathbf{x}_m)\, \mathrm{d}\mathbf{x}, \tag{2}$$

or simply $v_k = b * (u \circ \tau_k)$.

The PSF $b : \Omega_k \to [0, 1]$ is a probability density function that can be turned into $b_k : \Gamma \to [0, 1]$ by the change of variable $\mathbf{x}' = \tau_k(\mathbf{x})$ so that

$$v_k(\mathbf{x}_m) = \int_{\Gamma} u(\mathbf{x}') b_k(\mathbf{x}' - \tau_k(\mathbf{x}_m))\, \mathrm{d}\mathbf{x}' \tag{3}$$

There are several ways of computing the warped PSF b_k, depending on how we model the initial PSF b. The more common assumption is to consider that b is a 2D Gaussian centered at position $\mathbf{x}_m$. Since Gaussian filters have infinite support, it is quite difficult to implement in practice. A simpler model of the PSF is to assume pixels are square and uniformly sensitive to light, so that the PSF is a uniform square density function. Noting A the area of a pixel centered on $(0, 0)$ in a source view k, we get

$$b(x, y) = \begin{cases} \frac{1}{A^2} & \text{if } -\frac{1}{A} \leq x, y \leq \frac{1}{A} \\ 0 & \text{elsewhere.} \end{cases} \tag{4}$$

Under the assumption that the warp τ_k is locally linear, the warped PSF is a uniformly distributed parallelogram. In this case, we can make an even stronger assumption by supposing that the warp preserve the pixels (their area and squared shape), which is actually untrue but largely simplify the implementation. From now on, we take a unit pixel area. Since the intensity is constant and equals $u(\mathbf{p})$ over all the pixel area in the target view, the previous convolution can be written as in [14]:

$$v_k(\mathbf{x}_m) = \sum_{\mathbf{p} \in \Gamma} u(\mathbf{p}) \int_{\mathbf{p}} b_k(\mathbf{x}' - \tau_k(\mathbf{x}_m))\, \mathrm{d}\mathbf{x}', \tag{5}$$

so the pixel intensity $\mathbf{m}$ in the source image is

$$v_k(\mathbf{m}) = \sum_{\mathbf{p} \in \Gamma} B_{k,m,p} u(\mathbf{p}), \tag{6}$$

where $B_{k,m,p} = \int_{\mathbf{p}} b_k(\mathbf{x}' - \tau_k(\mathbf{x}_m))\, \mathrm{d}\mathbf{x}'$ is the area of intersection between the projection of the pixel in the target view and a pixel $\mathbf{p}$ of this view. It is equivalent to bilinearly interpolate the intensities of u.

Fig. 2: Warp and blending weight discontinuities cause artifacts (in red). Left: a close-up to a view rendered by minimizing the energy 7. Right: a warp that presents visibility discontinuities that caused the artifacts.

3.2. Maximum A Posteriori (MAP) Estimation

The goal of the variational approach is to estimate the high resolution image u from the data $(v_k^*)_{k \in [1..K]}$, where K is the number of inputs views. The estimator of u maximizes the posterior which is the probability of finding u given the input data. One can show that this is equivalent to minimizing the energy

$$E(u) = E_{\text{intensity}}(u) + \lambda E_{\text{prior}}(u), \tag{7}$$

E_{prior}, often referred to as the *smoothness term*, comes from the image prior and prevents the emergence of high frequencies. λ is a parameter that controls the smoothness of the final solution. In this work, we use a total variation prior [12], $E_{\text{prior}}(u) = \int_\Gamma |\nabla u|$, which has several advantages over other more complicated image priors: this approach preserves strong edges and image contours, and is convex. A proof of convergence is given by Chambolle [4].

$E_{\text{intensity}}$, often referred as the *data term*, is derived from the likelihood given the input image intensity [24]. This accounts for how well the current solution fits the data in the intensity domain:

$$E_{\text{intensity}}(u) = \sum_{k=1}^{K} \frac{1}{2} \int_{\Omega_k} \omega_k(u)((b * (u \circ \tau_k) - v_k^*))^2\, \mathrm{d}\mathbf{x}. \tag{8}$$

The terms $\omega_k(u)$ are the per-pixel contribution of each input view k. They depend on the gradient of the current solution u and the geometric uncertainty of the 3D reconstruction.

3.3. Appending the Gradient Term

Since the geometry and the visibility may be discontinuous, the term $\omega_k(u)$ in 8 may be discontinuous too, resulting in artifacts in the synthesized image that may appear as spurious edges or textures (Fig. 2). The IBR method should prevent these contours from appearing: actually, an image contour synthesized in the target image should also be present in source images where this part of the scene is visible.

To enforce this property, we add an extra term $E_{gradient}(u)$ to the previous energy (8) that forces the current solution to also fit the data in the gradient domain:

$$\begin{aligned} E_{\text{gradient}}(u) &= -\log \mathrm{p}(\nabla v_0 \dots \nabla v_{K-1} | \nabla u) && (9)\\ &= -\sum_{k=0}^{K-1} \log \mathrm{p}(\nabla v_k | \nabla u) && (10)\\ &= \int_{\Omega_k} (\nabla v_k - \nabla v_k^*)^2 \, \mathrm{d}\mathbf{x} && (11)\\ &= \int_{\Omega_k} (\nabla (b * (u \circ \tau_k)) - \nabla v_k^*)^2 \, \mathrm{d}\mathbf{x} && (12) \end{aligned}$$

Finding u that minimizes this energy is equivalent to solving the Laplace equation:

$$\Delta((b * (u \circ \tau_k) - v_k^*) = 0, \tag{13}$$

where $\Delta = \nabla.\nabla$ denotes the Laplacian. We instantly deduce the derivative of the functional:

$$\mathrm{d}E_{\text{gradient}}(u) = (|\frac{\partial \tau_k}{\partial z}|^{-1} \, \bar{b} * (\Delta(b * (u \circ \tau_k)) - \Delta v_k^*)) \circ \beta_k. \tag{14}$$

The β_k are the backward warp that appear thank to the change of variable in the integral. $\bar{b}$ is the adjoint of the blur kernel b. The warps τ_k are those which were estimated beforehand, and thus lack precision. This uncertainty has a drastic effect on the computation of $\Delta(b*(u\circ\tau_k))$. As a consequence, we chose to compute the Laplacian of u first, then warp it in the Ω_k domain. Under the assumption that the warps τ_k can be locally linear, we neglect their second order derivatives and obtain:

$$\Delta(b*(u\circ\tau_k)) = b*\left(\frac{\partial \tau_k}{\partial x}^\intercal H_u \frac{\partial \tau_k}{\partial x} + \frac{\partial \tau_k}{\partial y}^\intercal H_u \frac{\partial \tau_k}{\partial y} \right). \tag{15}$$

$H_u = \frac{\partial \nabla u}{\partial \mathbf{x}}$ is the Hessian matrix of u. Due to uncertain depth maps that cause strong discontinuities in warps, the Hessian matrix may be very unstable. For the computation, and in this case only, we assume $\tau_k(\mathbf{x}) \approx \mathbf{x} + \mathbf{d}$ so that

$$\Delta(b*(u\circ\tau_k)) = b*(\mathrm{trace}(H_u)\circ\tau_k) = b*(\Delta u \circ \tau_k). \tag{16}$$

The final form of the energy to minimize is thus:

$$E(u) = \alpha E_{intensity}(u) + \gamma E_{gradient}(u) + \lambda E_{prior}(u). \tag{17}$$

We minimize the functional (17) via Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [2].

3.4. Discretization

One can wonder how we go from a continuous model of the problem to a numerical solution. For each pixel $\mathbf{m}$ of each input view k we obtain an equation similar to (6). Let $\mathbf{V}^*$ be the vector of all the pixels of every input view put in one column $(v_0(0) \quad v_0(1) \quad \dots \quad v_{K-1}(M-1))^\intercal$, $\mathbf{U}$ the column vector storing the current solution $(u(0) \quad \dots \quad u(N-1))^\intercal$, and $\mathbf{B}$ the $KM \times N$ matrix that stores the $B_{k,m,p}$ coefficients, we may naturally write $\mathbf{V} = \mathbf{BU}$. Consequently, we can express the energy (8) as a linear system:

$$E_{intensity}(\mathbf{U}) = (\mathbf{BU} - \mathbf{V}^*)^\intercal \mathbf{W} (\mathbf{BU} - \mathbf{V}^*), \tag{18}$$

where $\mathbf{W}$ is a $KM \times KM$ diagonal matrix that stores the weights $|J_{\mathbf{x}'}(\beta_k)|\omega_k$. To minimize this energy we derive the linear system, and obtain the normal equations that provide an estimator for the solution $\hat{\mathbf{U}}$:

$$\mathbf{B}^\intercal \mathbf{W} \mathbf{B} \hat{\mathbf{U}} = \mathbf{B}^\intercal \mathbf{W} \mathbf{V}^*. \tag{19}$$

The matrix $\mathbf{B}^\intercal\mathbf{WB}$ is generally not invertible. This linear system can be solved by any linear least square solver.

Akin to the data term on color, the data term on image gradient is

$$E_{grad}(\mathbf{U}) = (\mathbf{B}\nabla\mathbf{U} - \nabla\mathbf{V}^*)^\intercal (\mathbf{B}\nabla\mathbf{U} - \nabla\mathbf{V}^*) \tag{20}$$

and can be derived likewise.

4. EXPERIMENTS AND VALIDATION

4.1. 3D Reconstruction

Our method takes as input an unstructured set of source views with no particular structure [3], as opposed to view interpolation methods which usually take stereo pairs, or methods that are based on a structured light field [29]. To be as generic as possible, it only requires the warp functions τ_k that we obtain via a classic 3D reconstruction pipeline [9]. It consists in camera calibration via *bundle adjustment* [20], followed by an MVS reconstruction [10] to get a depth map for each input view (Fig. 3). Knowing the camera parameters and the depth of each input pixel, we can deduce per-pixels correspondences between the source images and the target

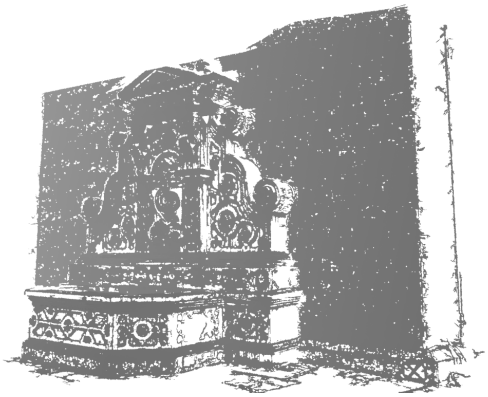


Fig. 3: The depth map of an input view.

Fig. 4: Results on datasets *fountain* and *herzjesu*. The bottom row shows some of the input views. Note that the parts of the target view that are not visible by any of the inputs are filled by a push/pull inpainting algorithm.

view. We use the depth to derive the blending weights as it is demonstrated in [24]. For further information about the reconstruction pipeline, more specifically occlusion handling and the derivation of the blending weights, see [21].

4.2. Results

A set of experiments (Figure 5) is performed on real views taken from Strecha's dataset [27], *fountain* and *herzjesu*. *Geometric proxies* are estimated according to the pipeline described in section 4.1. For both datasets we remove the central view, render it in the same dimensions as the input images and compare the result with the original for visual evaluation. All experiments were performed on GPU with an nVidia GTX Titan. Convergence is reached within 3 minutes for an input set of 11 input images of size 3072×2048. Since the convergence time strongly depends on the initialisation of the solution, the performance could be significantly increased. However, one drawback of our variational method is that it does not perform real-time rendering. To show that the quality of our results does not depend on the initialization, we start the optimization process from a null image.

Since some parts of the target view are not visible from the input views due to self-occlusion, inpainting is needed to fill potential holes. To that end, we implemented the *push/pull* algorithm as it is described in [13]. Firstly, the *push* stage decomposes the final image u into its Gaussian pyramid. At each level of the pyramid the image is filtered by a Gaussian-like 5×5 kernel and down-sampled by a factor 2. Only non-null value pixels contribute to create the upper-level image, so that at the coarsest level of the pyramid the image has no hole. Note that we do not need to down-sample to the 1×1 pixel image; in our experiments the 6×4 pixel image does not contain any holes. Secondly the *pull* stage propagates missing information from the top of the pyramid downwards to the finest level. At each level, holes are filled with the corresponding pixels from the upper coarser image.

At first we tested for a null gradient term ($\gamma = 0.0$). Since the linear system to be solved is ill-constrained, high frequencies appear in areas that few cameras see. These artifacts are emphasized by a very noisy depth estimation near occlusion regions (around the fish on the fountain or the Jesus on the wall). To remove these artifacts and eliminate high frequencies, a stronger smoothness term is commonly used. Consequently we increased the λ parameter that controls our *Total Variation* regularizer to 0.003. However the results are not that convincing: although most high frequencies are removed, some image features are lost compared to originals. To fix this problem, we achieved a third rendering where λ is set to its initial value (0.002) and the gradient data term is appended to the equation ($\gamma = 1.0$). For the final solution to keep the original colors of the input image, we keep the intensity data term as a bias but choose a small controlling parameter ($\alpha = 0.1$). One can notice that artifacts are completely removed, while preserving all of the image features. We conclude that appending the gradient data term prevents the emergence of spurious contours near visibility borders, hence guaranteeing the *continuity* property.

Some numerical results are presented in table 1. To show that our rendering algorithm outperforms state-of-the-art methods when synthesizing a specific view only, we generate several input views that we preliminarily removed from the input

	$\alpha = 1.0, \gamma = 0.0, \lambda = 0.002$ ([24, 29])	$\alpha = 1.0, \gamma = 0.0, \lambda = 0.003$ ([24, 29])	$\alpha = 0.1, \gamma = 1.0, \lambda = 0.002$ (ours)
fountain – view 2	21.03 132	21.09 120	**21.16** **107**
fountain – view 5	26.00 74	26.14 64	**26.36** **51**
fountain – view 8	22.00 140	22.08 125	**22.16** **111**
herzjesu – view 2	21.73 186	**21.96** 153	21.93 **143**
herzjesu – view 4	23.13 194	23.81 130	**23.90** **115**
herzjesu – view 6	18.08 349	18.26 287	**18.31** **273**

Table 1: Numerical results on real-world datasets [27]. Our method is compared against state-of-the-art methods [24, 29], for which there is no gradient constraints ($\gamma = 0.0$). For each result, the first value is the PSNR (bigger is better), the second value is DSSIM in units of 10^{-4}. $\mathrm{DSSIM} = 10^4(1 - \mathrm{SSIM})$ [28] (smaller is better). The best value is highlighted in bold. See text for a detailed description of the experiments.

set and kept as reference views for quantitative comparison. Two numerical measures are computed with respect to the reference view to evaluate our results: PSNR (the higher the better) and $\mathrm{DSSIM} = 10^4(1 - \mathrm{SSIM})$ (the lower the better). Although the increase of the PSNR score is noticeable but not conclusive, the strong improvement of the DSSIM demonstrates a higher structural similarity with the reference view, which accounts for the elimination of most visual artifacts.

5. CONCLUSION

We presented an image-based rendering method that renders a novel view from a generic and unstructured set of input views. This method is inspired by previous work by Pujades *et al* [24], which proposed a formulation for most of the "desirable properties" that were listed in the seminal work by Buehler *et al* [3], using a Bayesian formulation, and optimizing the target image in a variational framework. The only property that could not be formally derived was the *continuity* property, which states that the contribution of each input view to the pixels of the target image should be a continuous function of the pixel coordinates.

We showed that an alternative approach for enforcing the *continuity* property is to state that edges, contours or textures should not be created in the target image if they are not present in the source images. This results in an additional data term, based on image gradients, which can be added to the energy functional. The energy can then be solve by iteratively solving a linear system devised from the energy functional. The results show an improvement over previous intensity-based unstructured IBR methods, both in terms of objective image quality measurements, and in terms of subjective quality.

Limitations: Despite the noticeable improvement of the image quality, our method does not fully eliminate all the visible artifacts. It could be reworked to optimize directly the target image gradients, rather than intensities, and the target intensity could then be reconstructed by solving the Poisson equation, as is done in Kopf *et al* [15]. This should totally remove any variations in the synthesized image that come from the discontinuity of the visibility functions, and are still visible, although attenuated, in our results.

Future work: As shown in section 3, appending gradient constraints as a new energy data term is comparable to solving the Poisson equation. In the stitching literature [30], it is known as Laplacian blending. Considering that the Laplacian of an image behaves like a band-pass filter, blending the Laplacian of the input images is analogous to blending the images in the frequency domain for a specific band of frequencies that depends on the scale of the Laplacian. In our case the Laplacian is computed at the original scale of the image, level 0 of a Laplacian pyramid. Therefore we blend the images for the band of highest frequencies, and high-frequency artifacts due to naive intensity blending are prevented. Our future work will focus on an extension to all scales via the use of Laplacian pyramids, enabling a complete Laplacian blending to prevent not only high-frequency but also low-frequency artifacts.

6. REFERENCES

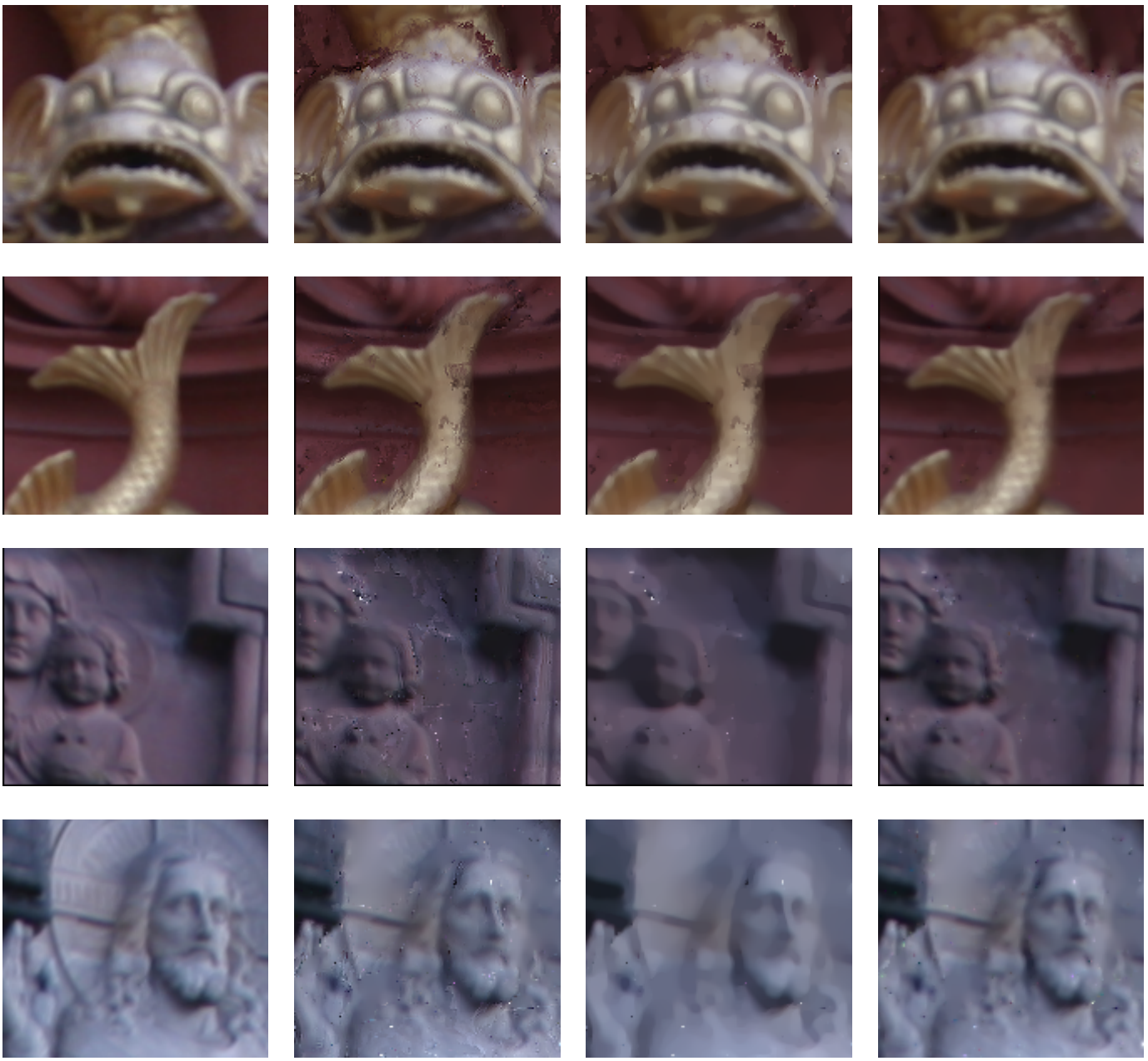
[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, September 2002.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In *IEEE ICASSP 2009*, pages 693–696, April 2009.

[3] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen. Unstructured lumigraph rendering. SIGGRAPH, pages 425–432, New York, NY, USA, 2001.

[4] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, January 2004.

[5] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis. Depth synthesis and local warps for plausible image-based navigation. *ACM Trans. Graph.*, 32(3):30:1–30:12, July 2013.



reference image	$\alpha = 1.0, \gamma = 0.0,$ $\lambda = 0.002$ [24, 29]	$\alpha = 1.0, \gamma = 0.0,$ $\lambda = 0.003$ [24, 29]	$\alpha = 0.1, \gamma = 1.0,$ $\lambda = 0.002$ ours

Fig. 5: Rendering the central view with different energy parameters. Each column shows the results on datasets *fountain* and *herzjesu* [27] by applying a special set of parameters $(\alpha, \gamma, \lambda)$ that control the amount of terms in the energy formula 17. Proposed approach is for $\gamma \neq 0$. State-of-the-art approach [24, 29] creates high-frequency artifacts due to blending intensities only. A higher smoothness parameter $\lambda = 0.003$ partially removes these artifacts at the cost of detail loss. Our approach preserves detail and show finer results by appending constraints of the gradients of the solution, forcing a Laplacian blending.

[6] A. Davis, M. Levoy, and F. Durand. Unstructured light fields. *Computer Graphics Forum*, 31(2pt1):305–314, 2012.

[7] A. Fitzgibbon, Y. Wexler, and A. Zisserman. Image-based rendering using image-based priors. In *IEEE ICCV, 2003. Proceedings*, pages 1176–1183 vol.2.

[8] J. Flynn, I. Neulander, J. Philbin, and N. Snavely. DeepStereo: Learning to predict new views from the world's imagery. 2015.

[9] S. Fuhrmann, F. Langguth, and M. Goesele. MVE – a multiview reconstruction environment. In *Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage (GCH)*, volume 6, page 8, 2014.

[10] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

[11] B. Goldluecke and D. Cremers. Superresolution texture maps for multiview reconstruction. In *2009 IEEE 12th ICCV*, pages 1677–1684, September 2009.

[12] B. Goldluecke and D. Cremers. An approach to vectorial total variation based on geometric measure theory. In *2010 IEEE CVPR*, pages 327–333, June 2010.

[13] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. SIGGRAPH '96, pages 43–54, New York, NY, USA, 1996. ACM.

[14] R. C. Hardie, K.J. Barnard, and E.E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6(12):1621–1633, December 1997.

[15] J. Kopf, F. Langguth, D. Scharstein, R. Szeliski, and M. Goesele. Image-based rendering in the gradient domain. *ACM Trans. Graph.*, 32(6):199:1–199:9, November 2013.

[16] A. Levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. In *Ninth IEEE ICCV, 2003. Proceedings*, pages 305–312 vol.1, October 2003.

[17] C. Lipski, F. Klose, and M. Magnor. Correspondence and depth-image based rendering a hybrid approach for free-viewpoint video. *IEEE Transactions on Circuits and Systems for Video Technology*, 24(6):942–951, June 2014.

[18] C. Lipski, C. Linz, K. Berger, A. Sellent, and M. Magnor. Virtual video camera: Image-based viewpoint navigation through space and time. *Computer Graphics Forum*, 29(8):2555–2568, 2010.

[19] J. McCann and N. S. Pollard. Real-time gradient-domain painting. SIGGRAPH '08, pages 93:1–93:7, New York, NY, USA, 2008. ACM.

[20] P. Moulon, P. Monasse, and R. Marlet. La bibliothèque openMVG: open source multiple view geometry. In *Orasis, Congrès des jeunes chercheurs en vision par ordinateur*, 2013.

[21] G. Nieto, F. Devernay, and J. Crowley. Rendu basé image avec contraintes sur les gradients. In *RFIA*, 2016.

[22] R. Ortiz-Cayon, A. Djelouah, and G. Drettakis. A bayesian approach for selective image-based rendering using superpixels. In *3D Vision (3DV), International Conference on*. IEEE, 2015.

[23] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 313–318, New York, NY, USA, 2003. ACM.

[24] S. Pujades, F. Devernay, and B. Goldluecke. Bayesian view synthesis and image-based rendering principles. In *2014 IEEE CVPR*, pages 3906–3913, June 2014.

[25] H. Y. Shum, S. C. Chan, and S. B. Kang. *Image-based rendering*. Springer Science & Business Media, 2008.

[26] S. N. Sinha, J. Kopf, M. Goesele, D. Scharstein, and R. Szeliski. Image-based rendering for scenes with reflections. *ACM Trans. Graph.*, 31(4):100, 2012.

[27] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *IEEE CVPR 2008*, pages 1–8, 2008.

[28] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, April 2004.

[29] S. Wanner and B. Goldluecke. Spatial and angular variational super-resolution of 4D light fields. In *Computer Vision – ECCV 2012*, number 7576 in Lecture Notes in Computer Science, pages 608–621. Springer Berlin Heidelberg, 2012.

[30] A. Zomet, A. Levin, S. Peleg, and Y. Weiss. Seamless image stitching by minimizing false edges. *IEEE Transactions on Image Processing*, 15(4):969–977, April 2006.